

The Changing Face of Computing

Sarita Adve

University of Illinois at Urbana-Champaign

sadve@illinois.edu

IIT-Bombay Techfest Lecture Series

This work is currently supported in part by DARPA, NSF, and by the Applications Driving Architecture (ADA) Research center (JUMP center co-sponsored by SRC and DARPA)

The Changing Face of Computing

- Research is the Foundation of Change
- The Changing Face: A Paradigm Shift in Computer Systems
- The Changing Face: Diversity and Inclusion

Who Invented the Iphone?



Processor: RISC, Berkeley/Stanford

Compiler: LLVM, Illinois

Battery: Lithium-ion, Texas

Internet, Multi-Touch Screen, GPS, ...

Inspired by: Peter Harsha et al.'s work on "Deconstructing the iPad: How Federally Supported Research Leads to Game-Changing Innovation," Task Force on American Innovation.

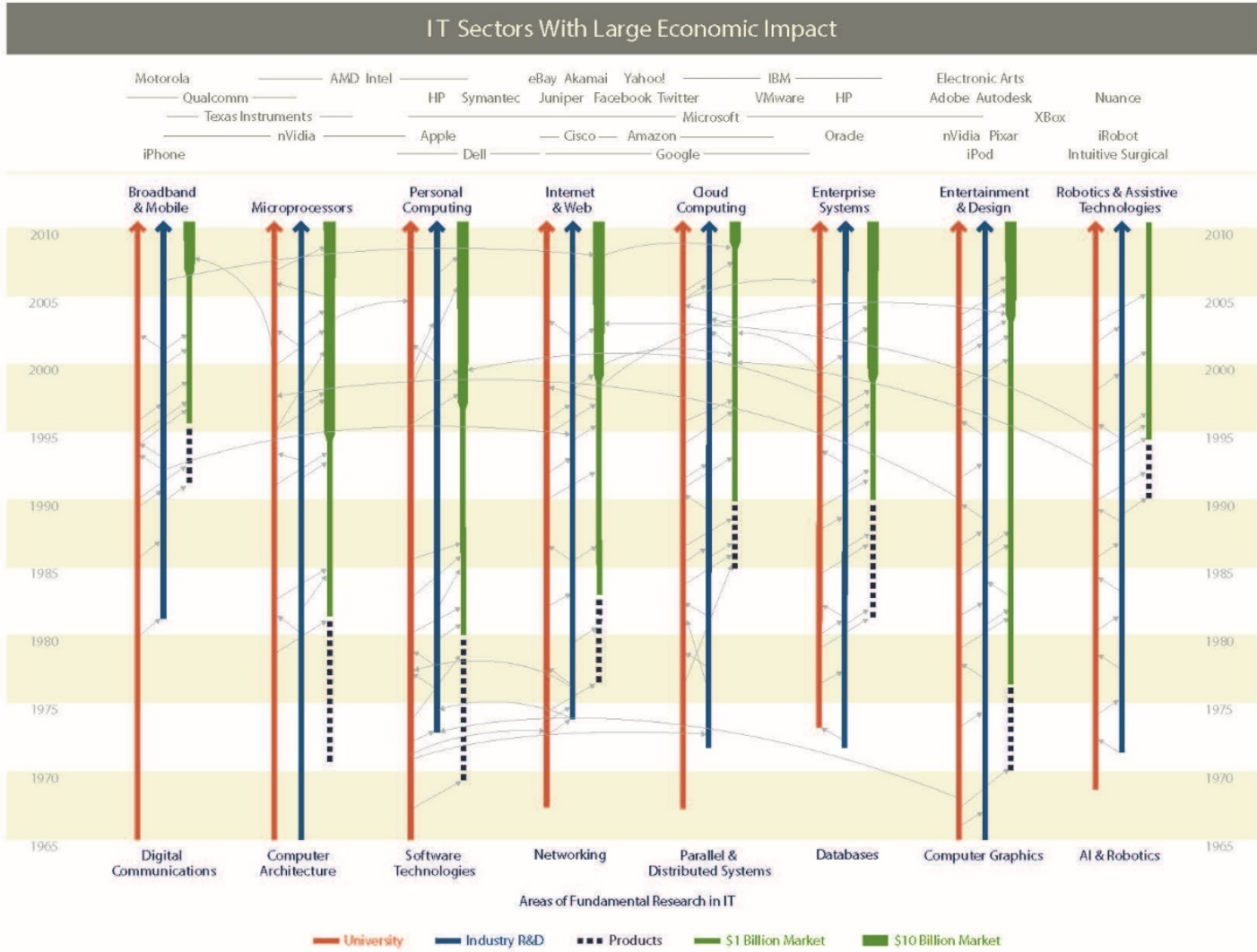


FIGURE I.1 Examples of the contributions of federally supported fundamental research to the creation of IT sectors, firms, and products with large economic impact. SOURCE: Reprinted from National Research Council, 2012, *Continuing Innovation in Information Technology*, The National Academies Press, Washington, D.C.

What is Research? How Does it Happen?

Pick a problem that excites you. Solve it.

- Question fundamentals
- Make an impact = Change minds.
- It takes time. Don't give up

My First Real Research Problem (1988)

[With my PhD advisor Mark Hill at Univ of Wisconsin]

Pick a problem that excites you. Solve it.

What is the best **memory consistency model** for a multiprocessor?

Question fundamentals

What is a memory consistency model? What is best?

Make an impact = Change minds.

Model adopted by popular programming languages: Java, C++, C, ...

It takes time. Don't give up.

First paper in 1990. Java adoption in 2005. C++ in 2008.

Impact continues on emerging systems

What is a Memory Model?

Initially X=Y=Flag=0

Thread 1

X = 26

Y = 90

...

Flag = 1

Thread 2

if (Flag == 1) {

... = Y ← 90

... = X ← 26

...

}

0

Memory model = What value can a read return?

Hardware/Software Interface:

Affects performance, programmability, portability

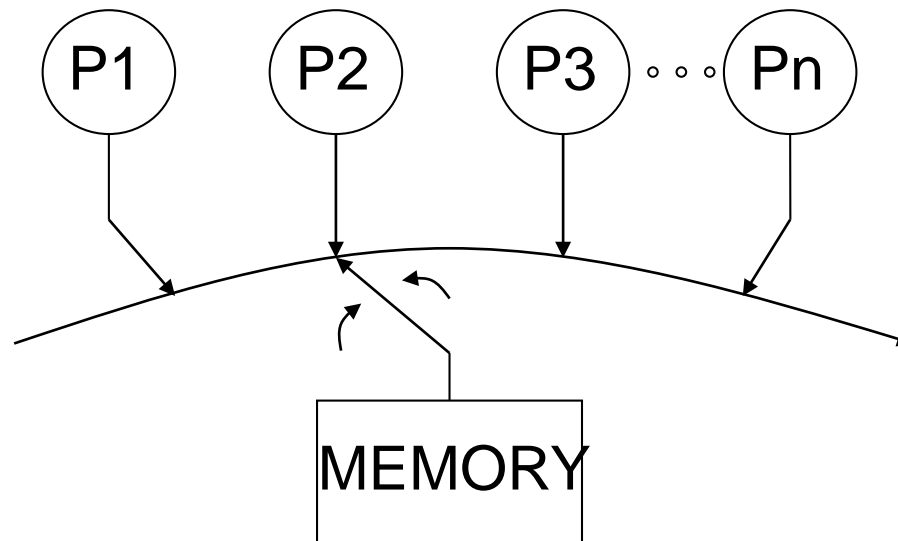
The Uniprocessor or Single Thread Model

- Program text defines total order = *program order*
- Uniprocessor model
 - Memory operations appear to execute one-at-a-time in program order
 - ⇒ Read returns value of last write
- BUT uniprocessor hardware
 - Overlap, reorder operations
- Model maintained as long as
 - Maintain control and data dependences
- ⇒ Easy to program + high performance

Implicit Multithreaded Memory Model

Sequential consistency (SC) [Lamport]

- Result of an execution appears as if
 - Memory operations of each thread in **program order**
 - All operations executed in some **sequential order**



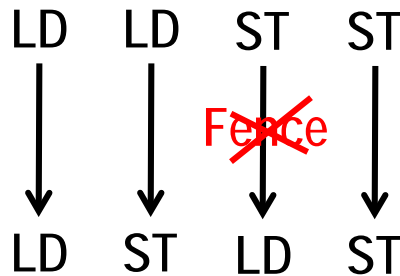
⇒ Easy to program BUT

Inhibits common optimizations in hardware and compiler

Prevalent View of What is a Memory Model

1988: Implementation/performance-centric view of memory model

- Order in which memory operations execute
- Different vendors w/ different orderings (models)
 - Alpha, Sun, x86, Itanium, IBM, AMD, HP, Cray, ...



- Complex, many ambiguities, ...

A new memory model virtually everyday

⇒ Higher performance BUT

Programmability? Portability?

What is a Memory Model?

Initially X=Y=Flag=0

Thread 1

X = 26

Y = 90

...

Flag = 1

Thread 2

if (Flag == 1) {

... = Y ← 90

... = X ← 26

...

}

0

Memory model = What value can a read return?

Hardware/Software Interface:

Affects performance AND programmability AND portability

An Alternate Programmer-Centric View

- One source of consensus
 - Programmers want SC to reason about programs
- But SC not practical today
 - How about the next best thing...

1989-93: An Alternate Programmer-Centric View

- Specify memory model as a contract
 - System gives sequential consistency
 - IF programmer obeys certain (easy) rules
- Key observation:
 - Programmers write well-synchronized or data-race-free programs even with SC
 - Data accesses can be optimized
- Data-race-free (DRF) model: SC for DRF programs [AdveHill90]
- Programmability AND Performance AND Portability

The Data-Race-Free (DRF) Model

- Data-race-free (DRF) model
 - Sequential consistency for data-race-free programs
 - Distinguish **data** vs. **synchronization (race)**
 - Data can be optimized \Rightarrow \uparrow performance for DRF programs

Initially $X=Y=Flag=0$

Thread 1

$X = 26$

$Y = 90$

...

$Flag = 1$

Thread 2

if ($Flag == 1$) {

... = Y

... = X

...

}

What is Research? How Does it Happen?

Pick a problem that excites you. Solve it.

- Question fundamentals
- Make an impact = Change minds.
- It takes time. Don't give up

Impact = Change Minds. Takes Time.

[with Bill Pugh, Jeremy Manson, Doug Lea, Hans Boehm, et al.]

2000-05: Java memory model [Manson, Pugh, Adve POPL'05]

– DRF model BUT racy programs need semantics

Problem: Incredibly hard to formalize a spec with racy programs without prohibiting common optimizations

Took five years to fix

2005: Java memory model: DRF + big mess

Impact = Change Minds. Takes Time.

[With Hans Boehm et al.]

2005-08: C++ memory model [Boehm, Adve PLDI'08]

- DRF model BUT need high performance; mismatched hardware
- Hardware vendors, software developers complained about DRF, but no option
- Compromise \Rightarrow DRF + big mess

Good news: After 20 years, convergence at last!

But: How to debug racy programs, how to avoid out of thin air values, no semantics for relaxed atomics, ...

CACM'10: Memory Models: A Case for Rethinking Parallel Languages and Hardware [Adve&Boehm]

Last Decade: Back to Fundamentals

[With Vikram Adve, Byn Choi, Rakesh Komuravelli, Matt Sinclair, Hyojin Sung]

- 2008-14: Software-centric view for coherence: DeNovo protocol
 - More performance-, energy-, and complexity-efficient than state-of-the-art
- BUT meanwhile: the end of Dennard and Moore's laws
 - Architecture enters golden age [Turing award 2018 speech]
 - *Déjà vu for coherence and consistency*

The Changing Face of Computing

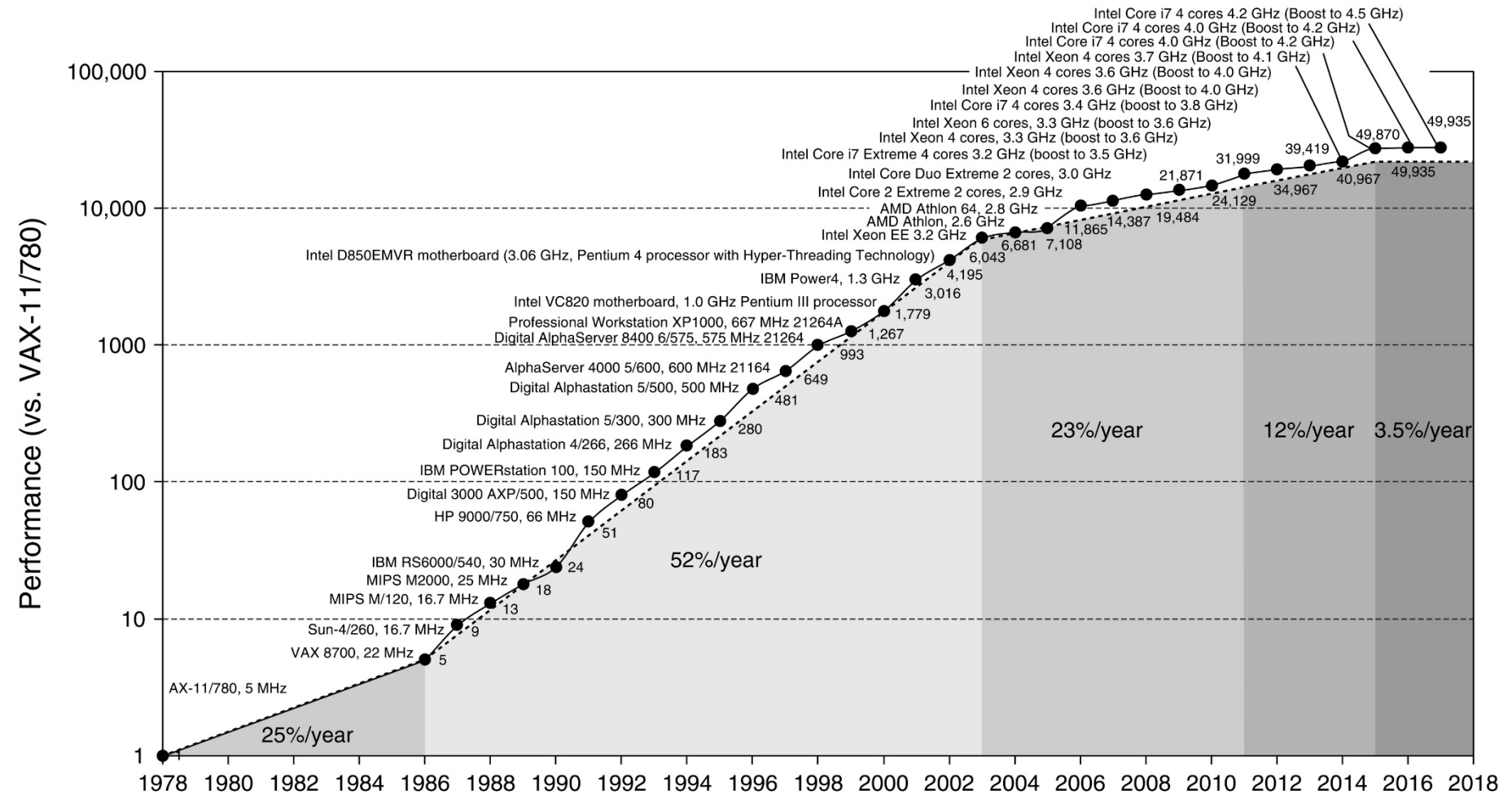
- Research is the Foundation of Change
- The Changing Face: A Paradigm Shift in Computer Systems
- The Changing Face: Diversity and Inclusion

Dennard's Law: Power

As transistors scale, power density stays constant

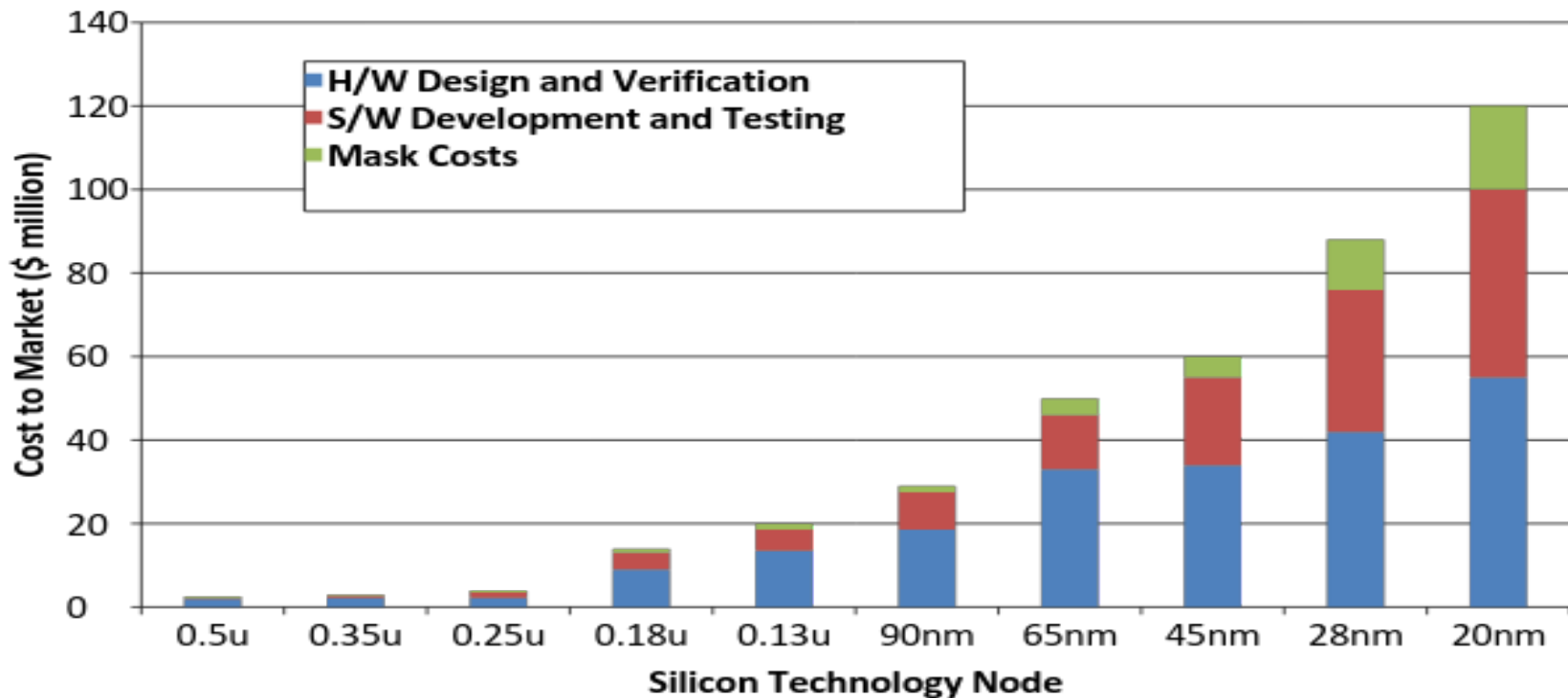
Cramming more transistors in same area does not increase power

Moore's Law: Processor Performance



Source: Hennessy and Patterson, Computer Architecture: A Quantitative Approach, 6th Edition
 © 2019 Elsevier Inc. All rights reserved.

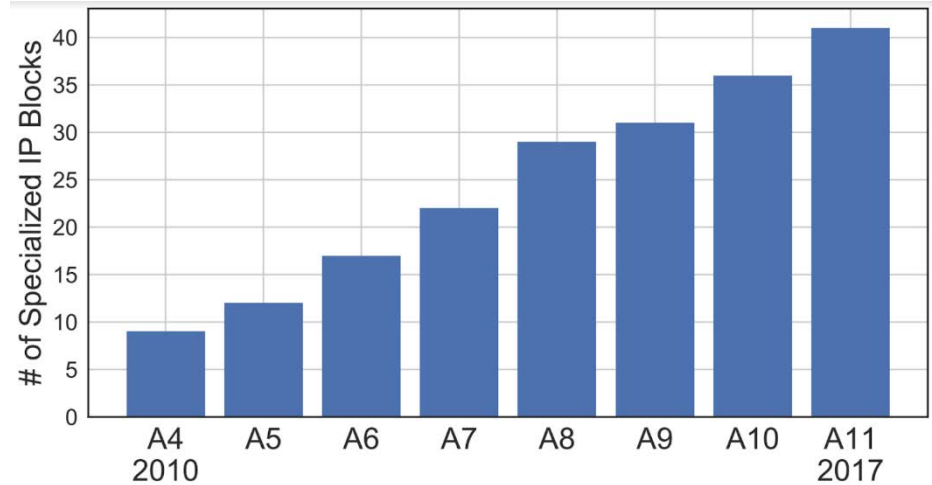
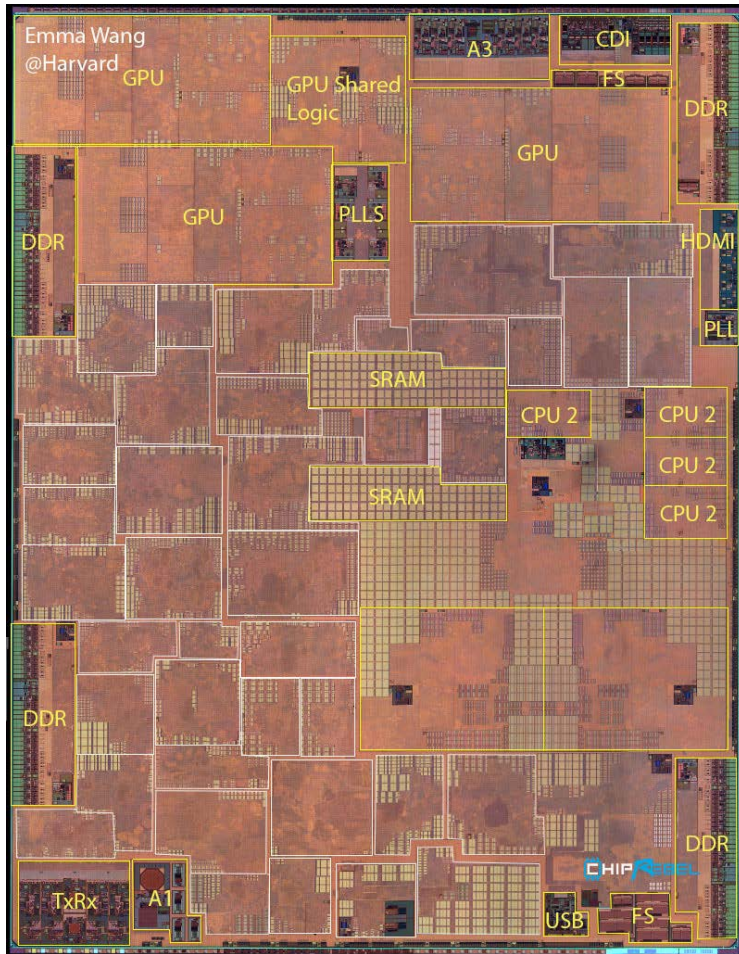
Hardware Cost to Market



Source: *International Business Strategies*

Graph from Todd Austin's seminar @ UIUC, 8/17

Specialization & (Parallel) Heterogeneity



Explosion of accelerators in SoCs

How to build specialized hardware?
What is the *hardware-software interface*?
How to build the software stack?

Source: Brooks, Wei group, <http://vlsiarch.eecs.harvard.edu/accelerators/die-photo-analysis>

Golden Age for Architecture (Systems) Research

- Need new abstractions
- Vertically integrated systems and projects
 - From application to hardware: design methodologies
- Emergent functional and performance criteria
 - Communication, verification, security, ...
- Technology opportunities
 - 2.5 and 3D scaling, In memory compute, Non-volatile memory, ...
- Open source hardware will change innovation ecosystem
- Push from above: New applications – ML, VR, robotics, IoT
 - Parallel, real-time, edge constrained, approximate

Needs research. Needs YOU!

A Riddle

A young boy and his father are on their way home when a distracted car driver hits them head-on. The father dies immediately. The boy is still alive and is transported in an ambulance to the hospital, where he is taken immediately into surgery.

However, the awaiting surgeon steps out of the operating room and says, "Please call Dr. Shah. I can't operate on this boy. He's my son!"

The question: Who is the surgeon?

The Changing Face of Computing

- Research is the Foundation of Change
- The Changing Face: A Paradigm Shift in Computer Systems
- The Changing Face: Diversity and Inclusion

A Riddle

A young boy and his father are on their way home when a distracted car driver hits them head-on. The father dies immediately. The boy is still alive and is transported in an ambulance to the hospital, where he is taken immediately into

surgery

We are all prone to bias against those different from us

Unless we consciously watch and correct for our unconscious bias

However, the awaiting surgeon steps out of the operating room and says, "Please call Dr. Shah. I can't operate on this boy. He's my son!"

The question: Who is the surgeon?

Bias Harms Diversity and Inclusion

- Diversity relates to gender, geography, economic status, caste, religion, ...
- My focus today is gender diversity

Why is Diversity Important

It's the right thing to do

But it is also the wise thing to do

Many studies show that diverse teams are more creative, innovative, and financially outperform non-diverse teams

(Two examples next)

Computing is going to continue to change lives, need to bring half the population to the decision making table

Why is Diversity Important

McKinsey&Company, "Delivering through Diversity," January 2018:
Leadership roles matter. Companies in the top-quartile for gender diversity on executive teams were 21% more likely to outperform on profitability and 27% more likely to have superior value creation. The highest-performing companies on both profitability and diversity had more women in line (i.e., typically revenue-generating) roles than in staff roles on their executive teams.

Nature editorial, "Science benefits from diversity," 6 June 2018:
A more representative workforce is more likely to pursue questions and problems that go beyond the narrow slice of humanity that much of science (biomedical science in particular) is currently set up to serve. Widening the focus is essential if publicly funded research is to protect and preserve its mandate to work to improve society. For example, a high proportion of the research that comes out of the Western world uses tissue and blood from white individuals to screen drugs and therapies for a diverse population. Yet it is well known that people from different ethnic groups can have different susceptibility to some diseases.

What's the Status in India?

Quick google search on “india gender diversity tech:”

<http://blog.belong.co/gender-diversity-indian-tech-companies>

Indian technology industry has 26% women in engineering

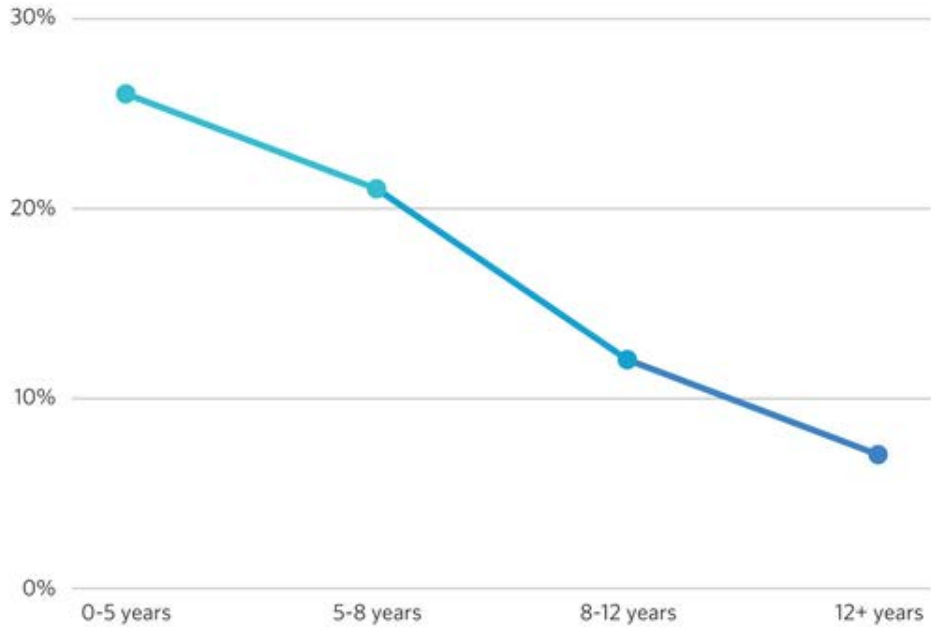
Men move into managerial roles faster

Nearly 50% women engineers quit tech

Only 7% reach the C-suite

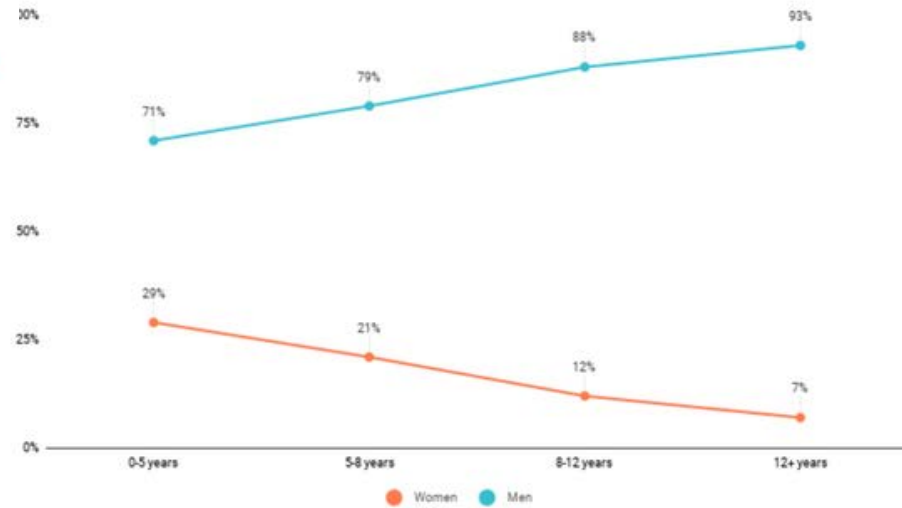
Cont...

How the percentage of women in tech drops over their career span



belong.co

How the gender gap widens over the years



Source: <http://blog.belong.co/gender-diversity-indian-tech-companies>

belong.co

Data from Academia - IITB

- IIT Bombay: 2001 to 2016 B.Tech. women were 5% to 9% (Source: Chayanika Shah, <http://www.fundamatics.net/women-in-iitbombay/>, 4/18)
- Distinguished alumni awards: 5/182, 2.7% (above source)
- Techfest Lecture series: 2/23 speakers are women
- The facebook post for my talk referred to me as MRS. Sarita Adve
 - My name is DR. Sarita Adve or PROF. Sarita Adve
 - Mrs. only draws attention to the fact that I am married which has no relevance to why I was invited to speak here

Data From Worldwide Research

ACM Fellow: No woman from South Asia until 2010

ACM doctoral dissertation award: No woman for last 16 years

IEEE CS: 7% of fellows are women as of July 2018 (didn't look for geography)

My field for computer architecture: First woman to win lifetime career award in 2018

First woman to win Turing award in 2006 (since 1966)

Why and What can YOU Do?

- Why?
 - Bias
 - Harassment
 - Cultural norms
- What can YOU (mostly for men) do?
 - Bias
 - Harassment
 - Cultural norms

The Changing Face of Computing

- Research is the Foundation of Change
- The Changing Face: A Paradigm Shift in Computer Systems
- The Changing Face: Diversity and Inclusion